Understanding the activity of contributors to VGI projects. How, why, where, and when do they contribute geographic information?

Peter Mooney

Geotechnologies Research Group, Department of Computer Science,
National University of Ireland Maynooth (NUIM), Co. Kildare, Ireland.
Email: peter.mooney@nuim.ie

Abstract. The crowdsourced collection of geographic information (Volunteered Geographic Information VGI) has moved to the center of the research agenda in Web technologies and GIS. Much of the current VGI research is tightly coupled with issues related to the quality of the collected geodata and possible conflation or comparison with other official sources of geodata. This paper presents research which attempts to quantitatively understand the activity patterns of citizen contributors to VGI projects. We argue that the concept of a large unbounded crowd of citizen contributors is something of a misnomer and that majority of work carried out in VGI collection and management is performed by a small, potentially, unconnected 'crowd' of contributors. We attempt to analyse how, why, where , and when these contributors work with VGI projects. OpenStreetMap is used as a case-study VGI project.

Keywords: VGI, Contributors, User Generated Content

1. Introduction and Motivation

Since Goodchild's now famous 'Citizen Sensors' paper (Goodchild, 2007) in 2007 there has been much research, discourse, and excitement generated by the concept of crowdsourcing geospatial data and information. The near ubiquity of smartphone devices and the appearance of 'always Internet connected' society have been important catalyst in the growth, variety, and availability of crowdsourced geospatial data. Examples abound with the most well known examples including: OpenStreetMap, Google Waze, Google Map Maker, Wikimapia, geo-coded Twitter streams, geo-coded photographs on services such as Flickr, etc.

There is no doubt that this Volunteered Geographic Information (VGI), which is a special case of the larger Web phenomenon known as user-generated content, is an exciting new paradigm in GIS and geospatial data production and distribution. Some authors have seen this rise in VGI as an inevitable direction and part of the 'evolution' of web-based GIS (Mooney and Corcoran 2013). However there are still major concerns within the GIS community and the geospatial data industry related to the quality, security and reliability of data generated and collected by VGI projects (Elwood and Leszczynski, 2011). There have been many studies carried out which have attempted to quantify the quality of geospatial data collected by VGI projects such as (Ludwig et al, 2012, Craglia et al , 2012, Zielstra et al 2013). For the most part these studies have generally positive results. It is now broadly accepted that there are recurring issues within VGI data such as: saturation coverage of urban areas whilst there is sparse coverage of rural areas, issues with semantics and metadata, geometrical accuracy, etc. However, against the backdrop of these issues, crowdsourced VGI data shows genuine strength in it's temporal accuracy (VGI tends to update very quickly) and how it is performing pioneering mapping work in socially and economically disadvantaged areas (for example the Map Kibera project in Kenya (Hodson, 2013)). The data and information generated for VGI projects should be considered fairly in terms of it's “fitness for purpose” and “fitness for use”. This means than rather than dismissing VGI out of hand there are many many applications where it can be of tremendous value and real-world use.

2. Overview of our presentation at ICA 2013

In our presentation at ICA 2013 we will argue that there has not been sufficient research attention directed to understanding “the crowd” in VGI. As mentioned above studies such as (Dobson 2013, Ludwig et al, 2012, Zeilstra and Hochmair, 2012) have shown that VGI data from OpenStreetMap is of very high quality in comparison to official data sources and usable for task applications such as navigation. But how much do we know about “the crowd” which generated this VGI? Our recent work (Mooney and Corcoran 2013a) shows that a small percentage of large crowdsourced communities (of up to 5,000 contributors) perform over 80% of all data collection and data editing in OSM for several cities around the world. This is illustrated in Figure 1 where the overall amount of work done by the top 10% and top 5% most prolific and most frequent contributors in each city is shown. This work in addition to more recent research will allow

us to provide quantitative and qualitative results for the following questions:

1. How big is the “crowd” in crowdsourced VGI projects?
2. Who are the 'senior' contributors (Neis and Zipf, 2012) – the contributors who carry out the majority of the work?
3. What areas are these contributors providing data for?
4. What is the spatial and temporal characteristics of the data contributed by these 'senior' contributors?

Our case-study dataset is the OpenStreetMap database. We will also highlight how these results are useful with the outputs generated from the first round of Internships funded by the 2012/2013 cooperation between EuroSDR, AGILE, and ESRI Europe.

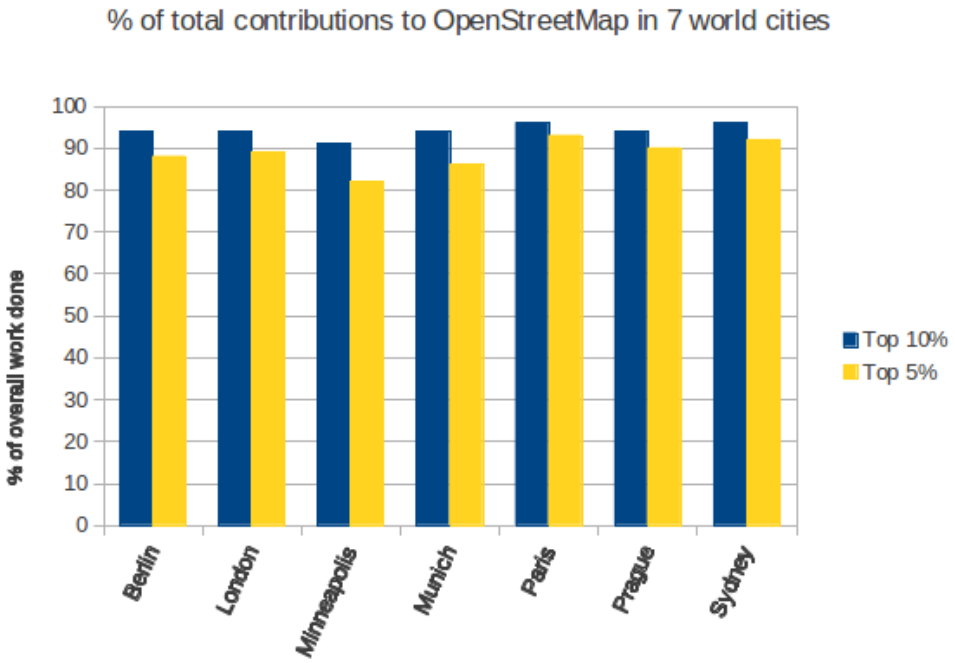

Figure 1: This figure shows the percentage of all contributions which were performed by the Top 10% and Top 5% most prolific contributors to OSM in a given city

3. Conclusions

We feel that gaining an improved understanding of the major or 'senior' will be very helpful in the overall evaluation of VGI generated by crowdsourced projects such as OpenStreetMap. Rather than seeing VGI data as the product of a largely chaotic and unregulated crowd of unknown 'citizen sensor' contributors we can now view the crowd as being comprised of a predominantly small group of highly ITC skilled VGI enthusiasts who are committed to the generation and open distribution of high quality spatial data. These group characteristics and dynamics are similar to those found in open source software projects (Mateos-Garcia and Steinmueller (2008), Giuri et al (2008)). Understanding who generates VGI, and how, where, when, and it is generated could potentially lead to greater trust and less uncertainty for end-users such as national mapping agencies, small and medium sized companies, and other industrial partners.

References

Craglia, M, Ostermann, F, and Spinsanti, L. Digital Earth from vision to practice: making sense of citizen-generated content. International Journal of Digital Earth Volume 5, Issue 5, 2012 pages 398-416

Elwood, S. and Leszczynski, A. Privacy reconsidered: New representations, data practices, and the geoweb. 2011 Geoforum 42(1): 6-15

Dobson, Michael W. VGI as a Compilation Tool for Navigation Map Databases. In Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice Sui, Daniel; Elwood, Sarah; Goodchild, Michael (Eds.) 2013, Springer. pp 307-327

Paola Giuri, Francesco Rullani, Salvatore Torrisi, Explaining leadership in virtual teams: The case of open source software, Information Economics and Policy, Volume 20, Issue 4, December 2008, Pages 305-315, http://dx.doi.org/10.1016/j.infoecopol.2008.06.002.

Goodchild, M.F. (2007). "Citizens as sensors: the world of volunteered geography". *GeoJournal* **69** (4): 211–221. doi:10.1007/s10708-007-9111-y.

Hodson, H. Citizen cartographers fill gaps in maps, New Scientist, Volume 219, Issue 2924, 6 July 2013, Page 20, http://dx.doi.org/10.1016/S0262-4079(13)61669-3.

Ludwig, Ina, and Voss, Agnes and Krause-Traudes, Maike. A Comparison of the Street Networks of Navteq and OSM in Germany In Advancing Geoinformation

Science for a Changing World Lecture Notes in Geoinformation and Cartography 2011, Springer pp 65-84

Mateos-Garcia, J, Steinmueller, W.E. The institutions of open source software: Examining the Debian community, Information Economics and Policy, Volume 20, Issue 4, December 2008, Pages 333-344,http://dx.doi.org/10.1016/j.infoecopol.2008.06.001.

Mooney, P, and Corcoran, P. Has OpenStreetMap a role in Digital Earth applications? International Journal of Digital Earth 2013, pp. 1-20, doi:10.1080/17538947.2013.781688

Mooney, P, and Corcoran P. Analysis of interaction and co-editing patterns amongst OpenStreetMap contributors, (Accepted), Transactions in GIS, July 2013a

Neis, P.; Zipf, A. Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap. *ISPRS Int. J. Geo-Inf.* **2012**, *1*, 146-165.

Zielstra, D., Hochmair, H. H., and Neis, P. (2013). Assessing the Effect of Data Imports on the Completeness of OpenStreetMap – A United States Case Study. Transactions in GIS, Volume 17, Issue 3, 315-334.

Zielstra, D., and Hochmair, H. H. (2012). Using free and proprietary data to compare shortest-path lengths for effective pedestrian routing in street networks. Transportation Research Record: Journal of the Transportation Research Board, 2299, 41-47.